

What is claimed is:

1. A document searching apparatus for
searching a document group having a link relation
5 for a document, comprising:

a link importance assigning unit weighting the
link relation and assigning link importance which
indicates importance of the document based on the
weighted link relation to the document; and
10 an accessing unit accessing the document based
on the link importance.

2. The document searching apparatus as set
forth in claim 1,
15 wherein said link importance assigning unit
includes:

a URL similarity calculating unit calculating
a URL similarity that is a similarity of URLs
(Uniform Resource Locators) that represent the
20 document,

wherein said link importance assigning unit
calculates the link importance based on the URL
similarity and the link relation of the document.

25 3. The document searching apparatus as set

forth in claim 1, further comprising:

a keyword extracting unit extracting text parts from the documents and extracting a keyword from the document contents.

5

4. The document searching apparatus as set forth in claim 3,

wherein said keyword extracting unit calculates an occurrence frequency of the keyword in the document, and

wherein said keyword extracting unit further comprises:

a keyword - document correlation calculating unit calculating the correlation of the keyword and the document based on the link importance and the occurrence frequency of the keyword.

5. The document searching apparatus as set forth in claim 4, further comprising:

a monitoring unit monitoring accesses from a user and generating an access log, and

wherein said keyword - document correlation calculating unit calculates the correlation based on the keyword occurrence frequency, the link importance, and the access log.

a document type determining unit determining a document type of the document based on the URL similarity, the number of links from the document, and the number of links to the document,

7. The document searching apparatus as set forth in one of claims 4, further comprising:

15 an index creating unit creating an index for accessing the document corresponding to pronunciation characters or spelling of the extracted keyword.

a selecting unit allowing the user to select a portion of the pronunciation characters or spelling of the keyword,

25 wherein said index creating unit places less

than a predetermined number of documents highly correlated with the keyword in the index based on the correlation calculated by said keyword - document correlation calculating unit, and

5 wherein said accessing unit accesses the document based on the selected keyword.

9. The document searching apparatus as set forth in one of claims 1, further comprising:

10 a collecting unit collecting the document from a network.

10. The document searching apparatus as set forth in claim 1,

15 wherein said link importance assigning unit causes the weight of the link relation between the documents with a high URL similarity to be decreased.

20 11. The document searching apparatus as set forth in claim 1,

 wherein said link importance assigning unit causes the document that is linked from important document and whose URL similarity is low to be
25 important.

12. The document searching apparatus as set forth in claim 1,

wherein said link importance assigning unit
5 causes the importance of document linked from many document whose URL similarity are high to be decreased.

13. The document searching apparatus as set forth in claim 1,

wherein the link importance of each document is defined as a solution of the following simultaneous linear equation (1), assuming that C_q is constant (the lower limit of the importance that
15 depends on each page) for each $p \in DOC$ and that when a page p is linked to a page q , the link weight $lw(p, q)$ is defined by the formula (2):

$$Wq = Cq + \sum_{p \in \text{Ref}(q)} Wp * lw(p, q) \quad \dots (1)$$

$$lw(p, q) = \text{diff}(p, q) / \sum_{i \in \text{Ref}(p)} \text{diff}(p, i) = \frac{1}{\text{sim}(p, q) \sum_{i \in \text{Ref}(p)} \frac{1}{\text{sim}(p, i)}} \quad \dots (2)$$

20 where $DOC = \{p1, p2, \dots, pN\}$ is a set of documents calculated for the link importance; Wp is

the link importance of the page p; Ref(p) is a set of pages linked from the page p; Refed(p) is a set of pages linking to the page p; sim(p, q) is the URL similarity of the pages p and q; diff(p, q) =
 5 1/sim(p, q) is the difference.

14. The document searching apparatus as set forth in one of claims 1,

wherein the URL similarity is determined based
 10 on characters of a URL containing a server address.

15. A document index creating apparatus for creating an index of a document group having a link relation, comprising:

15 a link importance assigning unit assigning a link importance to the document based on the link relation;

a keyword extracting unit extracting a keyword from the document;

20 an index creating unit creating an index for accessing the keyword based on pronunciation characters or spelling of the extracted keyword; and

an accessing unit accessing document assigned
 25 the link importance corresponding to the keyword

document type of the document based on a URL similarity representing the similarity of between URLs of the documents, the number of links to the document, and the number of links from the document,

5 wherein said index creating unit selects the document based on the document type and creates the link list of the selected document.

20. A document searching method for searching
10 a document group having a link relation for a document, comprising:

 assigning a link importance as an importance of the document calculated with weighting the link relation to the document; and

15 accessing the document based on the link importance.

21. The document searching method as set forth in claim 20, further comprising:

20 calculating a URL similarity that is a similarity of URLs (Uniform Resource Locators) that represent the document; and

 calculating the link importance based on the URL similarity and the link relation of the
25 document.

22. The document searching method as set forth in claim 20, further comprising:

extracting a keyword from the document.

5

23. The document searching method as set forth in claim 20, further comprising:

calculating an occurrence frequency of the keyword in the document, and

10 calculating the correlation of the keyword and the document based on the link importance and the occurrence frequency of the keyword.

24. The document searching method as set forth in claim 23, further comprising:

15 monitoring accesses from a user and generating an access log; and

calculates the correlation based on the keyword occurrence frequency, the link importance, 20 and the access log.

25. The document searching method as set forth in claim 23, further comprising:

determining a document type of the document 25 based on the URL similarity, the number of links to

the document, and the number of links from the document; and

selecting the document based on the document type and calculating the correlation of the
5 selected document.

26. The document searching method as set forth in one of claims 22, further comprising the step of:

10 creating an index for accessing the document corresponding to pronunciation characters or spelling of the extracted keyword.

27. The document searching method as set forth
15 in claim 26, further comprising the steps of:

placing less than a predetermined number of documents which are correlated with the keyword in the index;

selecting a portion of the pronunciation
20 characters or spelling of the keyword; and

accessing the document corresponding to the selected portion of the pronunciation characters or spelling of the selected keyword.

25 28. The document searching method as set forth

in one of claims 20, further comprising the step of:

collecting the document from a network.

5 29. A link list creating method for creating a link list for a document group having a link relation, comprising the steps of:

collecting the document from a network;

10 assigning a link importance which indicates importance of the document to the document based on the link relation;

determining a URL having a particular characteristic of a character string from the URLs of each document; and

15 creating a link list for listing less than a predetermined number of links to the document based on the link importance and the particular characteristic of the character string of the URL.

20 30. The link list creating method as set forth in claim 29, further comprising the steps of:

25 determining a document type of the document based on the URL similarity, the number of links to the document, and the number of links from the document; and

selecting the document based on the document
 type and creating the link list for the selected
 document based on the link importance and the
 particular characteristic of the character string
 5 of the URL.

31. A computer readable record medium for
 recording a program that causes a computer to
 execute a process for creating a link list for a
 10 document group having a link relation, the program
 comprising the steps of:

collecting documents from a network;
 assigning a link importance which indicates
 importance of the document to each document based
 15 on the link relation;

determining a URL having a particular
 characteristic of a character string from the URLs
 of documents; and

creating a link list for listing less than a
 20 predetermined number of links to the documents
 based on the link importance and the particular
 characteristic of the character string of the URL.